

Mumford-Shah and Potts Regularization for Manifold-Valued Data with Applications to DTI and Q-Ball Imaging

Andreas Weinmann, Laurent Demaret, Martin Storath^{*†}

October 8, 2014

Abstract

Mumford-Shah and Potts functionals are powerful variational models for regularization which are widely used in signal and image processing; typical applications are edge-preserving denoising and segmentation. Being both non-smooth and non-convex, they are computationally challenging even for scalar data. For manifold-valued data, the problem becomes even more involved since typical features of vector spaces are not available. In this paper, we propose algorithms for Mumford-Shah and for Potts regularization of manifold-valued signals and images. For the univariate problems, we derive solvers based on dynamic programming combined with (convex) optimization techniques for manifold-valued data. For the class of Cartan-Hadamard manifolds (which includes the data space in diffusion tensor imaging), we show that our algorithms compute global minimizers for any starting point. For the multivariate Mumford-Shah and Potts problems (for image regularization) we propose a splitting into suitable subproblems which we can solve exactly using the techniques developed for the corresponding univariate problems. Our method does not require any a priori restrictions on the edge set and we do not have to discretize the data space. We apply our method to diffusion tensor imaging (DTI) as well as Q-ball imaging. Using the DTI model, we obtain a segmentation of the corpus callosum.

^{*}Andreas Weinmann and Laurent Demaret are both with the Helmholtz Zentrum München, Germany. Martin Storath is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Switzerland.

[†]This work was supported by the German Federal Ministry for Education and Research under SysTec Grant 0315508. The first author acknowledges support by the Helmholtz Association within the young investigator group VH-NG-526. The third author was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 267439.

1 Introduction

In their seminal works [1, 2] Mumford and Shah introduced a powerful variational approach for image regularization. It consists of the minimization of an energy functional given by

$$\min_{u, C} \gamma |C| + \frac{\alpha}{q} \int_{\Omega \setminus C} |Du(x)|^q dx + \frac{1}{p} \int_{\Omega} d(u(x), f(x))^p dx. \quad (1)$$

Here, f represents the data and u is the target variable to optimize for. In the scalar case, u and f are real-valued functions on a domain $\Omega \subset \mathbb{R}^2$, d is the Euclidean metric, and Du denotes the gradient (in the weak sense). In contrast to Tikhonov-type priors, the Mumford-Shah prior penalizes the variation only on the complement of a discontinuity set C . Furthermore, the “length” $|C|$ (i.e., the outer one-dimensional Hausdorff measure) of this discontinuity set is penalized. The parameters $\gamma > 0$ and $\alpha > 0$ control the balance between the penalties. Basically, the resulting regularization is a smooth approximation to the image f which, at the same time, allows for sharp variations (“edges”) at the discontinuity set. The piecewise constant variant of (1) – often called Potts functional – corresponds to the degenerate case $\alpha = \infty$ which amounts to removing the second term in (1). Typical applications of these functionals are edge-preserving smoothing and image segmentation. For further information considering these problems from various perspectives (calculus of variation, stochastics, inverse problems) we exemplarily refer the reader to [3–12] and the references therein. These references also deal with theoretical questions such as, e.g., the existence of minimizers. Mumford-Shah and Potts problems are computationally challenging since one has to deal with non-smooth and non-convex functionals. Even for scalar data, both problems are NP-hard in dimensions higher than one [13–15]. This makes finding a (global) minimizer infeasible. However, due to its importance in image processing, many approximative strategies have been proposed for scalar- and vector valued data. Among these are graduated non-convexity [4], approximation by elliptic functionals [6], graph cuts [14], active contours [16], convex relaxations [17], and iterative thresholding approaches [10].

In recent years, regularization of manifold-valued data has gained a lot of interest. For example, sphere-valued data have been considered for SAR imaging [18] and non-flat models for color image processing [19–22]. Further examples are $SO(3)$ data expressing vehicle headings, aircraft orientations or camera positions [23], and motion group-valued data [24]. Related work dealing with the processing of manifold-valued data are wavelet-type multiscale transforms [23, 25, 26] and manifold-valued partial differential equations [27–29]; statistics on Riemannian manifolds are the topic of [30–36]. In medical imaging, a prominent example with manifold-valued data is diffusion tensor imaging (DTI). DTI allows to quantify the diffusional characteristics of a specimen non-invasively [37, 38]; see also the overview in [39]. DTI is helpful in the context of neurodegenerative pathologies such as schizophrenia [40, 41], autism [42] or Huntington’s disease [43]. In DTI, the data can be viewed as living in the Riemannian manifold of positive (definite) matrices; see, e.g., [44]. The underlying distance

corresponds to the Fisher-Rao metric [45] which is statistically motivated since the positive matrices (called diffusion tensors) represent covariance matrices. These tensors model the diffusivity of water molecules. Oriented diffusivity along fiber structures is reflected by the anisotropy of the corresponding tensors; typically, there is one large eigenvalue and the corresponding eigenvector yields the orientation of the fiber. In DTI, potential problems arise in areas where two or more fiber bundles are crossing because the tensors are not designed for the representation of multiple directions. In order to overcome this, the Q-ball imaging (QBI) approach [46–48] uses higher angular information to allow for multiple directional peaks at each voxel; it has been applied to diffusion tractography [49]. The Q-ball imaging data can be modeled by a probability density on the $3D$ -unit sphere called orientation distribution function (ODF). The corresponding space of ODFs can be endowed with a Riemannian manifold structure [50].

In the context of DTI, Wang and Vemuri consider a Chan-Vese model for manifold-valued data (which is a variant of the Potts model for the case of two segments) and a piecewise smooth analogue [51, 52]. Their method is based on a level-set active-contour approach which iteratively evolves the jump set followed by an update of the mean values (or a smoothing step for the piecewise smooth analogue) on each of the two segments. In order to reduce the computational load in their algorithms (caused by Riemannian mean computations for a very large amount of points) the authors resort to non-Riemannian distance measures in [51, 52]. Recently, a fast recursive strategy for computing the Riemannian mean has been proposed and applied to the piecewise constant Chan-Vese model in [53]. Related segmentation methods are K -means clustering [54], geometric flows [55] or level set methods [56, 57].

In this work, we propose algorithms for Mumford-Shah and Potts regularization for Riemannian manifolds (which includes DTI with the Fisher-Rao metric) for both signals and images. For manifold-valued data, the distance d in (1) becomes the Riemannian distance and the differential D can be understood in the sense of metric differentials [58]. For univariate Mumford-Shah and Potts problems, we derive solvers based on a combination of dynamic programming techniques developed in [2, 7, 59, 60] and proximal point splitting algorithms for manifold-valued data developed by the authors in [61]. Our algorithms are applicable for manifolds whose Riemannian exponential mapping and its inverse can be evaluated in reasonable time. For Cartan-Hadamard manifolds (which includes the manifold in DTI) our algorithms compute global minimizers for all input data. (We note that the univariate problems are not NP hard.) These results actually generalize to the more general class of Hadamard spaces. For Mumford-Shah and Potts problems for manifold-valued images (where the problems become NP-hard), we propose a novel splitting approach. Starting from a finite difference discretization of (1) we use a penalty method to split the problems into computationally tractable subproblems. These subproblems are closely related to univariate Mumford-Shah and Potts problems and can also be solved using the methods we developed for these problems in this paper. We note that our methods neither require a priori knowledge on the number of

segments nor a discretization of the manifold. We demonstrate the capabilities of our methods by applying them to two medical imaging modalities: DTI and Q-ball imaging. For DTI, we first consider several synthetic examples corrupted by Rician noise and show our algorithms potential for edge-preserving denoising. As specific medical imaging application, we obtain a segmentation of the corpus callosum for real human brain data. We conclude with experiments for Q-ball imaging.

1.1 Organization of the article

Section 2 deals with algorithms for the univariate Potts and Mumford-Shah problems for manifold-valued data. We start by presenting a dynamic programming approach for the univariate Potts and Mumford-Shah problem in Section 2.1. Then we use this approach to derive an algorithm for univariate Potts functionals for manifold-valued data in Section 2.2 and to derive an algorithm for the univariate Mumford-Shah problem in Section 2.3. An analysis of the derived algorithms is given in Section 2.4. In Section 3, we derive algorithms for the Potts and Mumford-Shah problems for manifold-valued images. We first deal with proper discretizations and then propose a suitable splitting into sub-problems that we solve using similar techniques as in the univariate case. We apply our algorithm to DTI data in Section 4 and to Q-ball data in Section 5.

2 Univariate Mumford-Shah and Potts functionals for manifold-valued data

In this section, we present solvers for Mumford-Shah and Potts problems for univariate manifold-valued data. These are not only important in their own right; variants of the derived solvers are also used as a basic building block for the proposed algorithm for the multivariate problems.

We first deal with some general issues; then, we derive the announced algorithms – first for the univariate Potts problem and then for the univariate Mumford-Shah problem; we conclude with an analysis of both algorithms.

In the univariate case, the discretization of the Mumford-Shah functional (1) and the Potts functional ($\alpha = \infty$ in (1)) is straightforward. The (equidistantly sampled) discrete Mumford-Shah functional reads

$$B_{\alpha,\gamma}(x) = \frac{1}{p} \sum_{i=1}^n d(x_i, f_i)^p + \frac{\alpha}{q} \sum_{i \notin \mathcal{J}(x)} d(x_i, x_{i+1})^q + \gamma |\mathcal{J}(x)|, \quad (2)$$

where d is the distance with respect to the Riemannian metric in the manifold M , $f \in M^n$ is the data, and \mathcal{J} is the jump set of x . The jump set is given by $\mathcal{J}(x) = \{i : 1 \leq i < n \text{ and } d(x_i, x_{i+1}) > s\}$ where the jump height s is related to the parameter γ via $\gamma = \alpha s^q / q$. Using a truncated power function we may

rewrite (2) in the Blake-Zisserman type form

$$B_{\alpha,s}(x) = \frac{1}{p} \sum_{i=1}^n d(x_i, f_i)^p + \frac{\alpha}{q} \sum_{i=1}^{n-1} \min(s^q, d(x_i, x_{i+1})^q), \quad (3)$$

where s is the argument the power function $t \mapsto t^q$ is truncated at.

The discrete univariate Potts functional for manifold-valued data reads

$$P_\gamma(x) = \frac{1}{p} \sum_{i=1}^n d(x_i, f_i)^p + \gamma |\mathcal{J}(x)|, \quad (4)$$

where d is the distance in the manifold and i belongs to the jump set of x if $x_i \neq x_{i+1}$.

We first of all show that the problems (2) and (4) have a minimizer. (We recall that certain variants of the continuous Mumford-Shah and Potts functional do not have a minimizer without additional assumptions; see, e.g., [11].)

Theorem 1. *In a complete Riemannian manifold the discrete Mumford-Shah functional (2) and the discrete Potts functional (4) have a minimizer.*

The proof is given in Appendix A. We note that the data spaces in applications are typically complete Riemannian manifolds.

2.1 The basic dynamic program for univariate Mumford-Shah and Potts problems

In order to find a minimizer of the Mumford-Shah problem (2) and the Potts problem (4), we use a general dynamic programming principle which was considered for the corresponding scalar and vectorial problems in various contexts; see, e.g., [2, 7, 59, 60, 62, 63]. We briefly recall the basic idea starting with the Mumford-Shah problem. It is convenient to use the notation

$$x_{l:r} = (x_l, \dots, x_r).$$

Assume that we have already computed minimizers x^l of the functional $B_{\alpha,\gamma}$ associated with the partial data $f_{1:l} = (f_1, \dots, f_l)$ for each $l = 1, \dots, r-1$ and some $r \leq n$. Then we compute x^r associated to data $f_{1:r}$ as follows. With each x^{l-1} of length $l-1$, we associate a candidate of the form $x^{l,r} = (x^{l-1}, h^{l,r}) \in M^r$ which is the concatenation of x^{l-1} with a vector $h^{l,r}$ of length $r-l+1$. This vector $h^{l,r}$ is a minimizer of the problem

$$\epsilon_{l,r} = \min_{h \in M^{r-l+1}} \sum_{i=l}^{r-1} \frac{\alpha}{p} d^q(h_i, h_{i+1}) + \frac{1}{p} \sum_{i=l}^r d^p(h_i, f_i), \quad (5)$$

and $\epsilon_{l,r}$ is the error of a best approximation on the (discrete) interval (l, \dots, r) . Then we calculate the quantity

$$\min_{l=1, \dots, r} \{B_{\alpha,\gamma}(x^{l-1}) + \gamma + \epsilon_{l,r}\}, \quad (6)$$

which we will see to coincide with the minimal functional value of $B_{\alpha,\gamma}$ for data $f_{1:r}$ (cf. Theorem 2 and Theorem 3). Then, we set $x^r = x^{l^*,r}$, where l^* is a minimizing argument in (6). We successively compute x^r for each $r = 1, \dots, n$ until we end up with full data f . Actually, only the l^* and the $\epsilon_{l,r}$ and not the vectors x^r have to be computed in this selection process; in a postprocessing step, the solution can be reconstructed from this information; see Algorithm 1 and [60] for further details. With these improvements, the dynamic programming skeleton (without the cost for computing the approximation errors $\epsilon_{l,r}$) has quadratic cost with respect to time and linear cost with respect to space. In practice, the computation can be accelerated significantly by pruning the search space [64, 65].

In order to adapt the dynamic program for the Potts problem (4) the only modification required is that the approximation errors on the intervals $\epsilon_{l,r}$ read

$$\epsilon_{l,r} = \min_{h \in M} \frac{1}{p} \sum_{i=l}^r d^p(h, f_i), \quad (7)$$

and the candidates are of the form $x^{l,r} = (x^{l-1}, h^{l,r})$, where $h^{l,r} \in M^{r-l+1}$ is constant and componentwise equals a minimizer h^* of (7) on the interval l, \dots, r . We next deal with the computation of these minimizers.

2.2 An algorithm for univariate Potts functionals for manifold-valued data

In order to make the dynamic program from Section (2.1) work for the Potts problem for manifold-valued data, we see from Section (2.1) that we have to compute the approximation errors $\epsilon_{l,r}$ given in (7) in the Riemannian manifold M . This means we are faced with the problem of computing a minimizer for the manifold-valued data $f_{l:r} = (f_l, \dots, f_r)$ and then to calculate the corresponding approximation error.

We first consider the case $p = 2$ which amounts to the “mean-variance” situation. Since our data live in a Riemannian manifold, the usual vector space operations to define the arithmetic mean are not available. However, it is well known (cf. [31, 44, 66, 67]) that a minimizer

$$z^* \in \arg \min_{z \in M} \sum_{i=1}^N d(z, z_i)^2 \quad (8)$$

is the appropriate definition of a mean $z^* \in \text{mean}(z_1, \dots, z_N)$ of the N elements z_i on the manifold M . A mean is in general not uniquely defined since the minimization problem has no unique solution in general. If the z_i are contained in a sufficiently small ball, however, the solution is unique. We then replace the “ \in ” symbol by an “=” symbol and call z^* the mean. The actual size of the ball where minimizers are unique depends on the sectional curvature of the manifold M ; for details and for further information we refer to [66, 67].

In contrast to the Euclidean case there is no closed form expression of the intrinsic mean defined by (8) in Riemannian manifolds. A widespread method for computing the intrinsic mean is the gradient descent approach (already mentioned in [66]) given by

$$z^{(k+1)} = \exp_{z^{(k)}} \sum_{i=1}^N \frac{1}{N} \exp_{z^{(k)}}^{-1} z_i. \quad (9)$$

(Recall that the points z_1, \dots, z_N are the points for which the intrinsic mean is computed.) Information on convergence related and other issues can, e.g., be found in the papers [31, 68] and the references therein. Newton’s method was also applied to this problem in the literature; see, e.g., [69]. It is reported in the literature and also confirmed by the authors’ experience that the gradient descent converges rather fast; in most cases, 5-10 iterations are enough. This might explain why this relatively simple method is widely used.

For general $p \neq 1$, the gradient descent approach works as well. The case $p = 1$ amounts to considering the intrinsic median and the intrinsic absolute deviation. In this case, the gradient descent (9) is replaced by a subgradient descent which in the differentiable part amounts to rescaling the tangent vector given on the right-hand side of (9) to length 1 and considering variable step sizes which are square-integrable but not integrable; see, e.g., [70].

A speedup using the structure of the dynamic program is obtained by initializing with previous output. More precisely, when starting the iteration of the mean for data $f_{l+1:r}$, we can use the already computed mean for the data $f_{l:r}$ as an initial guess. We notice that this guess typically becomes even better the more data items we have to compute the mean for, i.e., the bigger $r - l$ is. This is important since this case is the computational more expensive part and a good initial guess reduces the number of iterations needed.

A possible way to reduce the computation time further is to approximate the mean by a certain iterated two-point averaging construction (known as geodesic analogues in the subdivision context) as explained in [71]. Alternatively, one could use a “log – exp” construction (also known from subdivision; see [23]) which amounts to stopping the iteration (9) after one step.

The proposed algorithm for univariate Potts functionals for manifold-valued data is summarized in Algorithm 1.

2.3 An algorithm for univariate Mumford-Shah functionals for manifold-valued data

In order to make the dynamic program from Section 2.1 work for the Mumford-Shah problem with manifold-valued data, we have to compute the approximation errors $\epsilon_{l,r}$ in (5). To this end, we compute minimizers of the problem

$$V_\alpha(x; f) = \frac{1}{p} \sum_i d^p(x_i, f_i) + \alpha \frac{1}{q} \sum_i d^q(x_i, x_{i+1}). \quad (10)$$

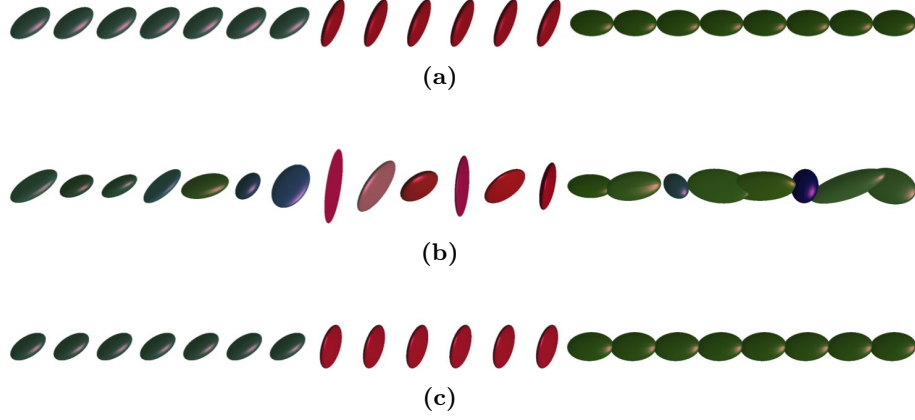


Figure 1: (a) Synthetic piecewise constant signal; (b) noisy data (Rician noise with $\kappa = 85$); (c) Potts regularization ($p, q = 1$) using Algorithm 1 with parameter $\gamma = 84.5$. The signal is reconstructed almost perfectly; the exact jump locations are obtained.

Here x is the target variable and f is the data. These are L^p - V^q type problems: the data term is a manifold ℓ^p distance and the second term is a q th variation; in particular, $q = 1$ corresponds to manifold-valued total variation. Solvers for these problems have been developed in the authors' paper [61]. We briefly recall the approach concentrating on the univariate case; for details we refer to [61]. We decompose the functional (10) into the sum $V_\alpha = F + \alpha \sum_i G_i$, where we let $G_i(x) = \frac{1}{q} d^q(x_i, x_{i+1})$ and $F(x) = \frac{1}{p} \sum_i d^p(x_i, f_i)$. For each of these summands, we can explicitly compute their proximal mappings defined by

$$\text{prox}_{\lambda G_i} x = \arg \min_y \left(\lambda G_i(y) + \frac{1}{2} d^2(x, y) \right). \quad (11)$$

They are given in terms of points on certain geodesics. In detail, we get

$$\begin{aligned} (\text{prox}_{\lambda G_i} x)_i &= [x_i, x_{i+1}]_t, \\ (\text{prox}_{\lambda G_i} x)_{i+1} &= [x_{i+1}, x_i]_t. \end{aligned} \quad (12)$$

where $[x, y]_t$ denotes the point reached after time t on the unit speed geodesic which is starting in x and going to y . For the practically relevant cases $q = 1, 2$ the parameter t has an explicit representation: for $q = 1$, we have $t = \lambda$, if $\lambda < \frac{1}{2} d(x_i, x_{i+1})$, and $d(x_i, x_{i+1})/2$ else; for $q = 2$ we get $t = \frac{\lambda}{1+2\lambda} d(x_i, x_{i+1})$. Similarly, the proximal mapping of F is given by

$$(\text{prox}_{\lambda F})_i(x) = [x_i, f_i]_s. \quad (13)$$

For $p = 1$, we have $s = \lambda$ if $\lambda < d(x_i, f_i)$, and $d(x_i, f_i)$ else; for $p = 2$, we obtain that $s = \frac{\lambda}{1+\lambda} d(x_i, f_i)$. We notice that the above proximal operators are uniquely defined if there is precisely one shortest geodesic joining the two points involved. Otherwise, one has to resort to set-valued mappings. Uniqueness is

given for the class of Cartan-Hadamard manifolds which includes the data space in DTI considered in Section 4.

Equipped with these proximal mappings we apply a cyclic proximal point algorithm for manifold-valued data [72]: we apply the proximal mappings of $F, \alpha G_r, \dots, \alpha G_l$ (with parameter λ) and iterate this procedure. During the iteration, we decrease the parameter λ_k in the k th iteration in a way such that $\sum_k \lambda_k = \infty$ and $\sum_k \lambda_k^2 < \infty$.

A speedup using the structure of the dynamic program is obtained by initializing with previous output as explained for the Potts problem in Section 2.2. The proposed algorithm for univariate Mumford-Shah functionals with manifold-valued data is summarized in Algorithm 1.

Algorithm 1: Algorithm for the Mumford-Shah problem (2) and the Potts problem (4) for univariate manifold-valued data

```

begin
  // Find optimal partition
   $B_0 \leftarrow -\gamma$ ;
  for  $r \leftarrow 1, \dots, n$  do
    for  $l \leftarrow 1, \dots, r$  do
      // Mumford-Shah case (Sec. 2.3):
       $\epsilon \leftarrow \min_{h \in M^{r-l+1}} V_\alpha(h; f_{l:r})$  // use Alg. of Sec. 2.3
      // Potts case (Sec. 2.2):
       $\epsilon \leftarrow \min_{h \in M} \sum_{i=l}^r d^p(h, f_i)$  // use Alg. of Sec. 2.2
       $b \leftarrow B_{l-1} + \gamma + \epsilon$ ;
      if  $b < B_r$  then
         $B_r \leftarrow b$ ;
         $p_r \leftarrow l - 1$ ;
      end
    end
  end
  // Reconstruct solution from partition
   $r \leftarrow n; l \leftarrow p_r$ ;
  while  $l > 0$  do
    // Mumford-Shah case (Sec. 2.3):
     $h^* \leftarrow \arg \min_{h \in M^{r-l+1}} V_\alpha(h; f_{l+1:r})$  // use Alg. of Sec. 2.3
    // Potts case (Sec. 2.2):
     $h' \leftarrow \arg \min_{h \in M} \sum_{i=l+1}^r d^p(h, f_i)$  // use Alg. of Sec. 2.2
     $h^* \leftarrow (h', \dots, h')$ ;
     $x_{l+1:r}^* \leftarrow h^*$ ;
     $r \leftarrow l; l \leftarrow p_r$ ;
  end
  return  $x^*$ 
end

```

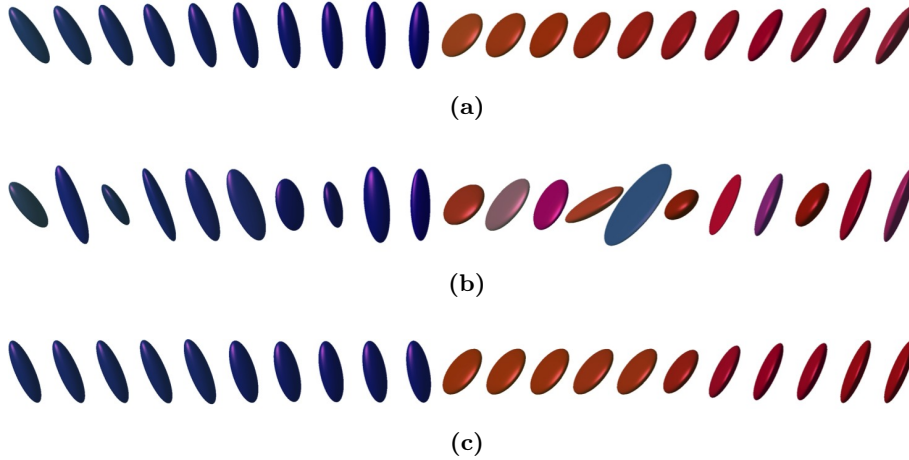


Figure 2: (a) Synthetic piecewise smooth signal; (b) noisy data (Rician noise with $\kappa = 70$); (c) Mumford-Shah regularization ($p, q = 1$) using Algorithm 1 with parameters $\alpha = 1.45$ and $\gamma = 1.5$. The noise is removed while preserving the jump.

2.4 Analysis of the univariate Potts and Mumford-Shah algorithms

We first obtain that our algorithms yield global minimizers for data in the class of Cartan-Hadamard manifolds which includes many symmetric spaces. Prominent examples are the spaces of positive matrices (which are the data space in diffusion tensor imaging) and the hyperbolic spaces. These are complete simply-connected Riemannian manifolds of nonpositive sectional curvature. For details we refer to [73] or to [74]. In particular, in these manifolds, geodesics always exist and are unique shortest paths.

Theorem 2. *In a Cartan-Hadamard manifold, Algorithm 1 produces a global minimizer for the univariate Mumford-Shah problem (2) (and the discrete Potts problem (4), accordingly).*

The proof is given in Appendix B.

We notice that this result generalizes to the more general class of (locally compact) Hadamard spaces. These are certain metric spaces generalizing the concept of Cartan-Hadamard manifolds; see, e.g., [75]. Examples of Hadamard spaces which are not Cartan-Hadamard manifolds are the metric trees in [75]. The validity of Theorem 2 for (locally compact) Hadamard spaces may be seen by inspecting the proof noticing that all steps rely only on features of these spaces.

For analysis of general complete Riemannian manifolds, we first notice that, in this case we have to deal with questions of well-definedness. We consider the Potts functional and data f_1, \dots, f_n . For each (discrete) subinterval $[l, r]$, a corresponding mean $h^{l,r}$ is defined as a minimizer of (8) for data f_l, \dots, f_r .

Although such a minimizer exists by the coercivity and continuity of the functional, it might not be unique. Furthermore, an algorithm such as gradient descent only computes a local minimizer for general input data. For data not too far apart, however, the gradient descent produces a global minimizer of (8) (since then the corresponding functional is convex). If data are so far apart that the operations in the manifold are not even well-defined it might be likely that they do not belong to the same segment. Hence, let us consider a constant C_K such that, if points belong to a C_K -ball with center in the compact set K , then their mean is uniquely defined and obtained by converging gradient descent. Assuming that the data lie in K , we call a partition of $[1, n]$ admissible if for any interval $[l, r]$ in this partition the corresponding data $f_{l:r}$ are centered in a common C_K -ball. We get the following result.

Theorem 3. *Let M be a complete Riemannian manifold. Then the univariate Potts problem given in Algorithm 1 with $p = 2$ produces a minimizer of the discrete Potts problem (4) when restricting the search space to candidates whose jump sets correspond to admissible partitions.*

The proof can be found in Appendix B. This result can be easily generalized to the general case $p \geq 1$.

3 Mumford-Shah and Potts problems for manifold-valued images

We now consider Mumford-Shah and Potts regularization for manifold-valued images. In contrast to the univariate case, finding global minimizers is not tractable anymore in general. In fact, the Mumford-Shah problem and the Potts problem are known to be NP hard in dimensions higher than one even for scalar data [13, 15]. Therefore, the goal is to derive approximative strategies that perform well in practice.

In the following it is convenient to use the notation $d^p(x, y)$ for the p -distance of two manifold-valued images $x, y \in M^{m \times n}$, i.e.

$$d^p(x, y) = \sum_{i,j} d^p(x_{ij}, y_{ij}).$$

We further define the penalty function

$$\Psi_a(x) = \sum_{i,j} \psi(x_{(i,j)+a}, x_{ij})$$

with respect to some finite difference vector $a \in \mathbb{Z}^2 \setminus \{0\}$. Here, we instantiate the potential function ψ in the Mumford-Shah case by

$$\psi(w, z) = \frac{1}{q} \min(s^q, d(w, z)^q). \quad (14)$$

and in the Potts case by

$$\psi(w, z) = \begin{cases} 1, & \text{if } w \neq z, \\ 0, & \text{if } w = z, \end{cases} \quad (15)$$

for $w, z \in M$.

In higher dimensions, the discretization of the Mumford-Shah and Potts problem is not as straightforward as in the univariate case. A simple finite difference discretization with respect to the coordinate directions is known to produce undesired block artifacts in the reconstruction [76]. The results improve significantly when including further finite differences such as the diagonal directions [65, 76, 77]. We here use a discretization of the general form

$$\min_{x \in M^{m \times n}} \frac{1}{p} d^p(x, f) + \alpha \sum_{s=1}^R \omega_s \Psi_{a_s}(x), \quad (16)$$

where the finite difference vectors $a_s \in \mathbb{Z}^2 \setminus \{0\}$ belong to a neighborhood system \mathcal{N} . The values $\omega_1, \dots, \omega_R$ are non-negative weights. We focus on the neighborhood system

$$\mathcal{N} = \{(1, 0); (0, 1); (1, 1); (1, -1)\}$$

with the weights $\omega_1 = \omega_2 = \sqrt{2} - 1$ and $\omega_3 = \omega_4 = 1 - \frac{\sqrt{2}}{2}$ as in [65]. For further neighborhood systems and weights we refer to [65, 76]. We next show the existence of minimizers of the discrete functional (16).

Theorem 4. *Let M be a complete Riemannian manifold. Then the discrete Mumford-Shah and Potts problems (16) both have a minimizer.*

The proof is given in Appendix A.

We next propose a splitting approach for the discrete Mumford-Shah and Potts problems. To this end, we rewrite (16) as the constrained problem

$$\begin{aligned} \min_{x_1, \dots, x_R} \quad & \sum_{s=1}^R \frac{1}{pR} d^p(x_s, f) + \alpha \omega_s \Psi_{a_s}(x_s) \\ \text{subject to } & x_s = x_{s+1} \text{ for all } 1 \leq s \leq R. \end{aligned} \quad (17)$$

Here, we use the convention $x_{R+1} = x_1$. (Note that x_1, \dots, x_R are $m \times n$ images.) We use a penalty method (see e.g. [78]) to include the constraints into the target functional and get the problem

$$\min_{x_1, \dots, x_R} \sum_{s=1}^R \omega_s p R \alpha \Psi_{a_s}(x_s) + d^p(x_s, f) + \mu_k d^p(x_s, x_{s+1}).$$

We use an increasing coupling sequence $(\mu_k)_k$ which fulfills the summability condition $\sum_k \mu_k^{-1/p} < \infty$. Optimization with respect to all variables simultaneously is still not tractable, but our specific splitting allows us to minimize

the functional blockwise, that is, with respect variables x_1, \dots, x_R separately. Performing the blockwise minimization we get the algorithm

$$\begin{cases} x_1^{k+1} \in \arg \min_x pR\omega_1\alpha\Psi_{a_1}(x) + d^p(x, f) + \mu_k d^p(x, x_R^k), \\ x_2^{k+1} \in \arg \min_x pR\omega_2\alpha\Psi_{a_2}(x) + d^p(x, f) + \mu_k d^p(x, x_1^{k+1}), \\ \vdots \\ x_R^{k+1} \in \arg \min_x pR\omega_R\alpha\Psi_{a_R}(x) + d^p(x, f) + \mu_k d^p(x, x_{R-1}^{k+1}). \end{cases} \quad (18)$$

We notice that each line of (18) decomposes into univariate subproblems of Mumford-Shah and Potts type, respectively. For example, we obtain

$$(x_1)_{:,j} \in \arg \min_{z \in M^n} pR\omega_1\alpha\Psi(z) + d^p(z, f_{:,j}) + \mu_k d^p(z, (x_R^k)_{:,j}) \quad (19)$$

for the direction $a_1 = (1, 0)$.

The subproblems are almost identical with the univariate problems of Section 2. Therefore, we can use the algorithms developed in Section 2 with the following minor modification. For the Potts problem, the approximation errors are now instantiated by

$$\epsilon_{l,r} = \min_{h \in M} \sum_{i=l}^r d^p(h, f_{ij}) + \mu_k d^p(h, (x_R^k)_{ij}),$$

for the subproblems with respect to direction a_1 (and analogously for the other directions a_2, \dots, a_R .) This quantity can be computed by the gradient descent explained in Section 2.2. In the Mumford-Shah case, we have

$$\epsilon_{l,r} = \min_{h \in M^{r-l+1}} \sum_{i=l}^{r-1} pR\omega_1\alpha d^q(h_i, h_{i+1}) + \sum_{i=l}^r d^p(h_i, f_{ij}) + \sum_{i=l}^m \mu_k d^p(h_i, (x_R^k)_{ij}).$$

The only difference to (5) is the extra “data term”

$$F'(h) = \sum_{i=l}^r d^p(h_i, (x_R^k)_{ij}).$$

Its proximal mapping has the same form as the proximal mapping of F in Section 2.3. Thus, we only need to complement the cyclic proximal point algorithm for the L^p - V^q problem of Section 2.3 by an evaluation of the proximal mapping with respect to F' .

We eventually show convergence.

Theorem 5. *For Cartan-Hadamard manifold-valued images the algorithm (18) for both the Mumford-Shah and the Potts problem converge.*

The proof is given in Appendix C.

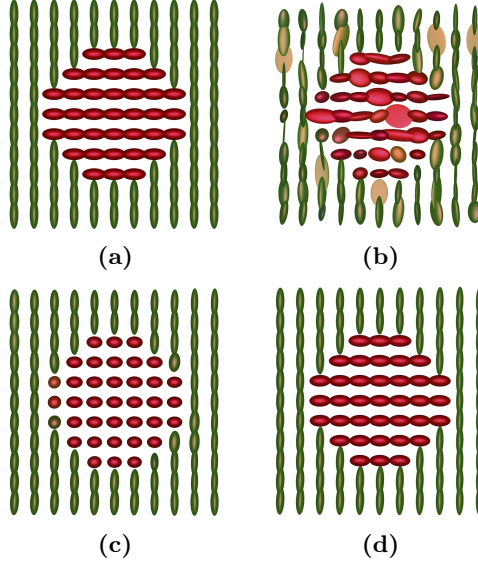


Figure 3: (a) Synthetic DT image, (b) noisy data (Rician noise of level 75), (c) L^1 -TV reconstruction (using TV parameter $\alpha = 0.65$), (d) Potts reconstruction ($p, q = 1$) with parameter $\gamma = 3.75$. While the L^1 -TV reconstruction decreases the contrast significantly, the Potts method yields an almost perfect reconstruction.

4 Application to Diffusion Tensor Images

The first application of our method is edge preserving denoising and segmentation of diffusion tensor images. Diffusion tensor imaging (DTI) is a non-invasive modality for medical imaging quantifying diffusional characteristics of a specimen. It is based on nuclear magnetic resonance [37, 38]. Prominent applications are the determination of fiber tract orientations [37], the detection of brain ischemia [79], and studies on autism [42], to mention only a few. Regularization of DT images is important in its own right and, in particular, serves as a processing step in many applications. It has been studied in a number of papers; we exemplarily mention [44, 51, 80, 81].

In DTI, the diffusivity of water molecules is encoded into a so-called diffusion tensor. This means that the data sitting in each pixel (or voxel) of a diffusion tensor image is a positive (definite symmetric) 3×3 matrix D . The space of positive matrices Pos_3 is a Riemannian manifold when equipped with the Riemannian metric

$$g_D(W, V) = \text{trace}(D^{-\frac{1}{2}} W D^{-\frac{1}{2}} V D^{-\frac{1}{2}}); \quad (20)$$

for details, see, e.g., [44]. Here the symmetric matrices W, V represent tangent vectors in the point D . Besides its mathematical properties, the practical advantage of the Riemannian metric (20) in comparison to the Euclidean metric is that it reduces the swelling effect ([27, 82]). On the flipside, the algorithms

and the corresponding theory become more involved.

4.1 Implementation of our algorithms for DTI

We now implement our algorithms for Mumford-Shah and Potts regularization for DTI data. Due to the generality of our algorithms, we only need an implementation of the Riemannian exponential mapping and its inverse to make them work on the concrete manifold. For the space of positive matrices, the Riemannian exponential mapping \exp_D is given by

$$\exp_D(W) = D^{\frac{1}{2}} \exp(D^{-\frac{1}{2}} W D^{-\frac{1}{2}}) D^{\frac{1}{2}}.$$

Here D is a positive matrix and the symmetric matrix W represents a tangent vector in D . The mapping \exp is the matrix exponential. The inverse of the Riemannian exponential mapping is given by

$$\exp_D^{-1}(E) = D^{\frac{1}{2}} \log(D^{-\frac{1}{2}} E D^{-\frac{1}{2}}) D^{\frac{1}{2}}.$$

for positive matrices D, E . The matrix logarithm \log is well-defined since the argument is a positive matrix. The matrix exponential and logarithm can be efficiently computed by diagonalizing the symmetric matrix under consideration and then applying the scalar exponential and logarithm functions to the eigenvalues. The distance between D and E is just the length of the tangent vector $\exp_D^{-1}(E)$ which can be explicitly calculated by $d(D, E) = (\sum_{l=1}^3 \log(\kappa_l)^2)^{\frac{1}{2}}$, where κ_l is the l^{th} eigenvalue of the matrix $D^{-\frac{1}{2}} E D^{-\frac{1}{2}}$.

The space of positive matrices becomes a Cartan-Hadamard manifold with the above Riemannian metric (20). Hence the theory developed in this paper fully applies; in particular, the univariate algorithms for DTI data produce global minimizers for all input data (see Theorem 2); furthermore, the algorithm (18) converges, and all its subproblems are solved exactly.

4.2 Synthetic data

The data measured in DTI are so-called diffusion weighted images (DWIs) D_v which capture the directional diffusivity in the direction v . The relation between the diffusion tensor image f and the DWIs D_v at some pixel p is given by the Stejskal-Tanner equation

$$D_v(p) = A_0 e^{-b \ v^T S(p) v}, \quad (21)$$

where $b, A_0 > 0$ are empirical parameters. For our simulation, we used $b = 800$ and $A_0 = 1000$. The tensor $S(p)$ is commonly derived from the DWIs via a least square fit using (21). In our experiments we visualize the diffusion tensors by the isosurfaces of the corresponding quadratic forms. More precisely, the ellipse representing the diffusion tensor $S(p)$ at pixel p are the points x fulfilling $(x - p)^T S(p) (x - p) = c$, for some $c > 0$.

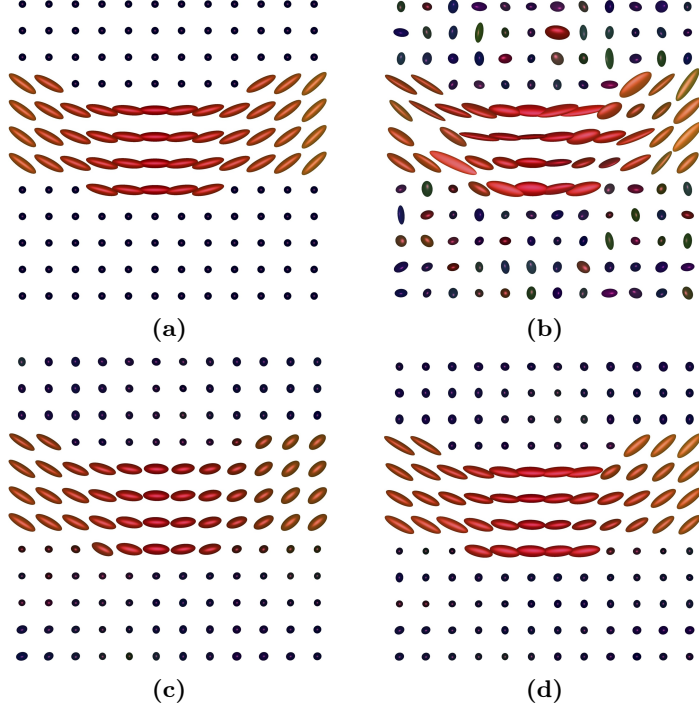


Figure 4: (a) Synthetic DT image, (b) noisy data (Rician noise of level 65), (c) L^1 -TV reconstruction (TV parameter $\alpha = 0.5$), (d) Mumford-Shah reconstruction ($p, q = 1$) using parameters $\gamma = 0.8$ and $\alpha = 5$. The noise is removed and the segments of the original image are recovered reliably.

We simulate noisy data using a Rician noise model [81, 84]. This means that we generate a noisy DWI $D'_v(p)$ by

$$D'_v(p) = \sqrt{(X + D_v(p))^2 + Y^2},$$

with clean data $D_v(p)$ and Gaussian variables $X, Y \sim N(0, \sigma^2)$. In our examples, we impose Rician noise to 15 diffusion weighted images and then compute the diffusion tensors according to the Stejskal-Tanner equation (21) using a least squares fit. We compare our results with L^p - V^q regularization, i.e., with minimizers of the two-dimensional analogue of (10) using the (globally convergent) cyclic proximal point algorithm of [61]. We optimized the model parameter with respect to the error to the groundtruth.

The univariate situation is illustrated in Fig. 1 for Potts and in Fig. 2 for Mumford-Shah regularization. Fig. 3 shows the effect of Potts regularization on a simple diffusion tensor image. The noise is removed and the segment boundaries are correctly recovered. The image in Fig. 4 possesses a certain variation within the segments. Therefore the (piecewise smooth) Mumford-Shah

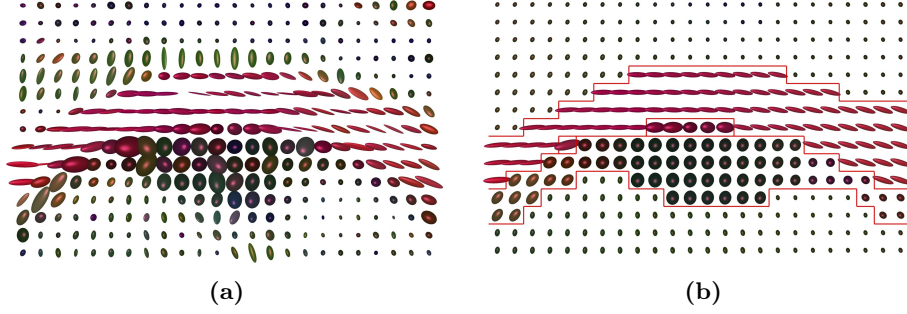


Figure 5: (a) Corpus callosum of a human brain from the Camino project [83]. (b) Mumford-Shah regularization ($p, q = 1$) using parameters $\alpha = 4.3$ and $\gamma = 2.9$. The noise is reduced significantly while the edges are preserved. In particular, the reconstruction induces a segmentation of the corpus callosum and its adjacent structures (red lines).

regularization is the proper method. As result, we obtain a piecewise smooth denoised image with preserved sharp edges.

4.3 Application to real data – segmentation of the corpus callosum

DTI is frequently used to study characteristics of the corpus callosum. The corpus callosum connects the right and the left hemisphere of the human brain. Typically, the first step of an analysis is the localization of the corpus callosum [42, 51]. We use our Mumford-Shah method for the segmentation of the corpus callosum of a human brain. This real data set stems from the Camino project [83]. In Fig. 5, we observe that our Mumford-Shah approach removes noise and preserves sharp boundaries between the oriented structures. In particular, the jump set yields an accurate segmentation of the corpus callosum.

5 Application to Q-Balls

In DTI the diffusion at each pixel/voxel is modeled via a single tensor. Typically, this tensor has one dominant eigenvalue with corresponding eigenvector pointing to the direction with maximal diffusivity. This direction is directly related with pathways of, e.g., neural fibers. DTI encounters difficulties for modeling voxels with intravoxel directional heterogeneity which, for example, occur at crossings of fiber bundles [85, 86]. In order to overcome these limitations, several approaches have been proposed [47, 85, 87, 88]. One of the most popular among these approaches is Q-ball imaging [86]. Here the tensor (seen as an ellipsoid parametrized over a ball) is replaced by a more general orientation distribution function (ODF) $\varphi : \mathbb{S}^2 \rightarrow \mathbb{R}$ where $\varphi(s)$ essentially corresponds to the diffusivity in direction s . Since the method allows for more flexibility, high angular resolution diffusion imaging (HARDI) data (see [46, 86]) are needed.

Further information can be found in the latter references.

5.1 The Q-ball manifold and the implementation of our algorithm for Q-ball imaging

In order to derive a Riemannian structure on the Q-ball manifold we follow the approach of [50]. The points in the (discrete) Q-ball manifold are “square-root parametrized” (discrete) ODFs which are a kind of samples of continuous ODFs $\varphi : \mathbb{S}^2 \rightarrow \mathbb{R}$ on a finite subset S of the sphere \mathbb{S}^2 with a preferably almost equidistant sampling. To be precise, a discrete ODF is a positive function $\varphi : S \rightarrow \mathbb{R}$ such that $\sum_{s \in S} \varphi^2(s) = 1$ (as proposed in [50]). Hence, a discrete ODF can be identified with a point on the sphere \mathbb{S}^{n-1} . Then the set Φ of all discrete ODFs is the intersection of the positive quadrant with the unit sphere in \mathbb{R}^n , and thus can be endowed with the Riemannian structure inherited from \mathbb{S}^{n-1} . Then the corresponding metric for the Q-ball manifold is given by

$$d(\varphi_1, \varphi_2) = \arccos \left(\sum_{s \in S} \varphi_1(s) \varphi_2(s) \right), \text{ for } \varphi_1, \varphi_2 \in \Phi.$$

The basic Riemannian operations have simple closed expressions. For a point φ on the unit sphere \mathbb{S}^{n-1} in \mathbb{R}^n and a non-zero tangent vector v to the sphere at φ , the exponential mapping is given by

$$\exp_{\varphi}(v) = \varphi \cdot \cos \|v\| + \frac{v \cdot \sin \|v\|}{\|v\|},$$

where $\|\cdot\|$ denotes the Euclidean norm in \mathbb{R}^n . The inverse of the exponential mapping is defined for any pair of points $\varphi_1, \varphi_2 \in \Phi$ by

$$\exp_{\varphi_1}^{-1}(\varphi_2) = d(\varphi_1, \varphi_2) \cdot \frac{\varphi_2 - \langle \varphi_1, \varphi_2 \rangle \varphi_1}{\|\varphi_2 - \langle \varphi_1, \varphi_2 \rangle \varphi_1\|}.$$

These explicit formulas for the Riemannian exp mapping and its inverse enable us to directly apply our algorithms for the regularization of Q-ball data.

5.2 Numerical experiments

We apply our algorithm to synthetic Q-ball data. Our examples simulate situations where two fiber bundles intersect. In the examples the size of the sampling set on the 2-sphere is $n = 181$ directions. In order to simulate noisy data, we use the method based on the so-called “soft equator approximation” [46]. We visualize a discrete ODF as a spherical polar plot. We compare our results with classical L^2 -Sobolev regularization (L^2 - V^2) using the cyclic proximal point algorithm of [61].

Our first example is a univariate signal (Fig. 6). It contains two kinds of Q-balls: one “tensor-like” with a single peak and another one with two peaks.

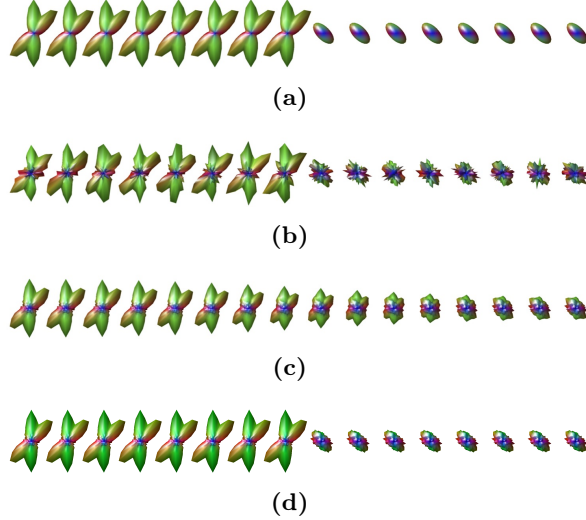


Figure 6: (a) Synthetic piecewise smooth Q-ball signal, (b) noisy data, (c) the manifold analogue of classical Sobolev regularization (L^2 - V^2 with $\alpha = 50$), (d) Mumford-Shah regularization ($p, q = 2$) with parameters $\alpha = 25, \gamma = 0.5$. Classical Sobolev regularization removes the noise, but it smooths out the jump; in contrast, the Mumford-Shah regularization removes the noise and preserves the jumps.

This illustrative example shows that, also in the Q-ball case, our regularization method removes the noise while preserving the jump and its location.

Our second experiment is a Q-ball valued image which simulates the crossing of two fiber bundles (Fig. 7). Here, we observe that our method removes the noise while preserving the fiber crossing and the directional structures encoded in the Q-balls as well as the edge structure in the image.

6 Conclusion and future research

In this paper, we proposed algorithms for the non-smooth and non-convex Mumford-Shah and Potts functionals for manifold-valued signals and images. We have shown the potential of our method by applying it to DTI and Q-ball imaging. Using the DTI model, we obtained a segmentation of the corpus callosum. For signals with values in Cartan-Hadamard manifolds (which includes the data space in diffusion tensor imaging), we have seen that our algorithms for univariate data produce global minimizers for any starting point. For the Mumford-Shah and Potts problems for image regularization (which is a NP hard problem) we have obtained convergence of the proposed splitting approach.

Topics of future research are the application of our algorithms to further nonlinear data spaces relevant for imaging. Another issue is to build a segmentation pipeline based on our method. Finally, from a theoretical side, it is interesting to further investigate convergence related questions for general

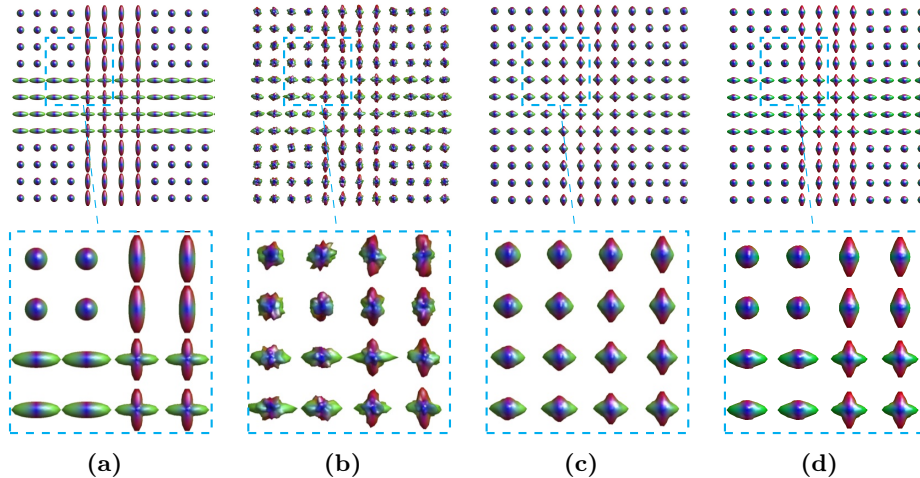


Figure 7: (a) Synthetic Q-Ball image, (b) noisy data, (c) the manifold analogue of classical Sobolev regularization (L^2 - V^2 with parameter $\alpha = 1$), (d) Mumford-Shah regularization ($p, q = 2$) with parameters $\alpha = 32, \gamma = 0.03$. The Sobolev regularization smooths out the edges and the crossing structures. The Mumford-Shah method recovers the edges as well as the crossings of the original image reliably.

Riemannian manifolds.

A Existence of minimizers

We supply the proofs of Theorem 4 and Theorem 1 which are statements on the existence of minimizers.

Proof of Theorem 4. We first show that the Mumford-Shah version of the discretization (16) has a minimizer. In the Mumford-Shah case, ψ is the truncated power function given by (14). Since ψ is continuous, so is Ψ_{a_s} for all s and therefore the whole functional given by (16) is continuous. On the other hand, the data term $d^p(x, f)$ is obviously coercive with respect to the Riemannian distance. This makes the overall functional coercive and confines points with small functional value to a bounded set. Since the manifold under consideration is complete, points with small functional value are confined to a compact set. Hence, the continuous functional takes its minimal value on this compact set and the corresponding point is a minimizer.

We come to the discrete Potts functional. Here we consider the discretization (16) where ψ is implemented by (15). With the same argument as for the Mumford-Shah functional above, the Potts functional is coercive with respect to the Riemannian distance. We show its lower semicontinuity. We have a look at Ψ_{a_s} which can be written as a sum of univariate jump functionals for manifold-valued data of the form $S : u \mapsto |\mathcal{J}(u)|$ from the Riemannian manifold M^j to the nonnegative integers (where j is the varying length of the data under

consideration.) If these functionals S were not lower semicontinuous, there would be a convergent sequence $u^n \rightarrow u$ with each $u^n \in M^j$ such that $|J(u)| > |J(u^n)|$ for sufficiently high indices n . Since $u^n \rightarrow u$ componentwise (with respect to the distance induced by the Riemannian metric), we get, using the triangle inequality, that

$$d(u_k^n; u_{k-1}^n) \rightarrow d(u_k; u_{k-1}).$$

This contradicts u having more jumps than u^n . Hence, the functionals S and, as a consequence, the functionals Ψ_{a_s} are lower semicontinuous. Using the continuity of the data term the discretization (16) of the Potts functional is lower semicontinuous. By its coercivity and the completeness of the manifold M , arguments with a small Potts value are located in a compact set. Hence, in the Potts case, (16) has a minimizer. This completes the proof. \square

Proof of Theorem 1. The assertion is a consequence of Theorem 4 when specifying to data defined on $\{1, \dots, n\} \times \{1\}$ choosing as single direction $a_1 = (1, 0)$. \square

B Univariate Mumford-Shah and Potts algorithms

We supply the proof of Theorem 2 which states that the algorithms proposed for the univariate problems produce global minimizers when the data live in a Cartan-Hadamard manifold.

Proof of Theorem 2. We start with the Mumford-Shah problem for manifold-valued data. For $l = 1, \dots, r$, we consider the first $l - 1$ data items $f_{1:l-1} = (f_1, \dots, f_{l-1})$. We let x^{l-1} be a minimizer of the corresponding functional $B_{\alpha, \gamma}^{l-1}$ for the truncated data $f_{1:l-1}$. Moreover, we let $h^{l,r} \in M^{r-l+1}$ be the result computed by our algorithm for the minimization of V_α according to Section 2.3 for data $f_{l:r}$. Since we are in a Cartan-Hadamard manifold, $h^{l,r}$ is a global minimizer of V_α by Theorem 2 in [61]. With each l we associate the candidate $x^{l,r} = (x^{l-1}, h^{l,r})$. On the other hand we consider an index l^* minimizing (6). We claim that the candidate $x^{l^*,r}$ is a minimizer of $B_{\alpha, \gamma}^r$. To see this, consider an arbitrary $x \in M^r$ and let k be its rightmost jump point k . If there is no such k , then x has no jumps and

$$B_{\alpha, \gamma}^r(x) = V_\alpha(x) \geq V_\alpha(x^{1,r}) \geq B_{\alpha, \gamma}^r(x^{l^*,r}).$$

The penultimate inequality is due to the fact that $x^{1,r}$ is a global minimizer of V_α in a Cartan-Hadamard manifold. The last inequality follows from (6). If k is the rightmost jump point of x , we have

$$B_{\alpha, \gamma}^r(x) = B_{\alpha, \gamma}^{k-1}(x) + \gamma + V_\alpha(x^{l,r}) \geq B_{\alpha, \gamma}^r(x^{l^*,r})$$

by (6). This shows the assertion of the theorem in the Mumford-Shah case using induction on r .

In the Potts functional case, we let x^{l-1} be a minimizer of the Potts functionals P_γ^{l-1} for the truncated data $f_{1:l-1}$. Then we let $h^{l,r} \in M^{r-l+1}$ be the result of the gradient (resp. subgradient) descent (9). Since we are in a Cartan-Hadamard manifold, $h^{l,r}$ agrees with the constant function on $[l, r]$ which is pointwise equal to the mean ($p = 2$), median ($p = 1$) or, in general, the minimizer of the right hand side of (7). Now we may proceed analogous to the Mumford-Shah case to conclude the assertion and complete the proof. \square

We proceed showing Theorem 3 which states that our algorithm yields a minimizer for the Potts problem when considering general complete Riemannian manifolds and candidates with admissible partitions.

Proof of Theorem 3. We use the notation of the proof of Theorem 2. Then, the x^{l-1} are minimizer of the corresponding Potts functionals P_γ^{l-1} for the truncated data $f_{1:l-1}$. (We notice that such a minimizer exists, since an interval consisting of one member is always admissible.) Furthermore, for admissible intervals $[l, r]$, $h^{l,r} \in M^{r-l+1}$ is pointwise equal to the computed Riemannian mean as explained in Section 2.2. The Riemannian mean minimizes the right hand side of (7). The candidates $x^{l,r} = (x^{l-1}, h^{l,r})$ and the minimizing index l^* are given as in the proof of Theorem 2 above. In order to show that $x^{l^*,r}$ is a minimizer, we consider an arbitrary $x \in M^r$ with an admissible partition. If x has no jump, then $P_\gamma(x) = \frac{1}{2} \sum_i d(x, f_i)^2 \geq P_\gamma(x^{1,r}) \geq P_\gamma(x^{l^*,r})$. Otherwise, let k be the rightmost jump point of x (which, by assumption, comes with an admissible partition). Then,

$$P_\gamma^r(x) = P_\gamma^{k-1}(x) + \gamma + V_\alpha(x^{l,r}) \geq P_\gamma^r(x^{l^*,r}).$$

which shows that $x^{l^*,r}$ is a minimizer. Now induction completes the proof. \square

C Mumford-Shah and Potts algorithms for images

We supply the proof of Theorem 5 stating that the algorithm in (18) converges in a Cartan-Hadamard manifold.

Proof of Theorem 5. We show that all iterates x_s^k converge to the same limit for all $s \in \{1, \dots, R\}$. Since we are in a Cartan-Hadamard manifold, x_1^{k+1} is a global minimizer of the functional $H_1(x) = pR\omega_1\alpha\Psi_{a_1}(x) + d^p(x, f) + \mu_k d^p(x, x_R^k)$ which is the first problem in (18). This follows by an argument similar to the proof of Theorem 2.

We have $H_1(x_1^{k+1}) \leq H_1(x_R^k)$ which means that

$$d^p(x_1^{k+1}, f) + \mu_k d^p(x_1^{k+1}, x_R^k) \leq pR\omega_1\alpha\Psi_{a_1}(x_R^k) + d^p(x_R^k, f). \quad (22)$$

In analogy, we get for the x_s^{k+1} , $s = 2, \dots, R$, using the other functionals in (18) that

$$d^p(x_s^{k+1}, f) + \mu_k d^p(x_s^{k+1}, x_{s-1}^k) \leq pR\omega_1\alpha\Psi_{a_s}(x_{s-1}^{k+1}) + d^p(x_{s-1}^{k+1}, f). \quad (23)$$

For both the Mumford-Shah and the Potts problem, the terms $\alpha\Psi_{a_1}(x_R^k)$ and $\alpha\Psi_{a_s}(x_{s-1}^{k+1})$, with $s = 2, \dots, R$, are uniformly bounded by a constant C which does not depend on k and s . This is because, for any input, $\alpha\Psi_{a_s}$ is bounded by αmn with the regularizing parameter α for the jump term of the functional under consideration, and m and n are the height and width of the image. Hence we can use (22) and (23) to get

$$\begin{aligned} d^p(x_1^{k+1}, x_R^k) &\leq \frac{C}{\mu_k} + \frac{1}{\mu_k}(d^p(x_R^k, f) - d^p(x_1^{k+1}, f)), \\ d^p(x_s^{k+1}, x_{s-1}^k) &\leq \frac{C}{\mu_k} + \frac{1}{\mu_k}(d^p(x_{s-1}^{k+1}, f) - d^p(x_s^{k+1}, f)). \end{aligned} \quad (24)$$

Now we may apply the inverse triangle inequality to the second summand on the right-hand side and get $d^p(x_R^k, f) - d^p(x_1^{k+1}, f) \leq d^p(x_R^k, x_1^{k+1})$. Then, a simple manipulation shows that

$$d^p(x_1^{k+1}, x_R^k) \leq \frac{C}{\mu_k - 1}, \quad d^p(x_s^{k+1}, x_{s-1}^k) \leq \frac{C}{\mu_k - 1}. \quad (25)$$

As a consequence, there is a constant D and an index k_0 such that, for all $k \geq k_0$,

$$d(x_R^{k+1}, x_R^k) \leq D\mu_k^{-\frac{1}{p}}.$$

Hence,

$$d(x_R^{k+1}, x_R^{k_0}) \leq D \sum_{l=k_0+1}^{k+1} \mu_l^{-\frac{1}{p}} < \infty,$$

and so the sequence x_R^{k+1} converges. By (24), the iterates x_s^k converge to the same limit for all $s = 1, \dots, R-1$. This completes the proof. \square

References

- [1] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 17, pages 137–154, 1985.
- [2] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [3] R.B. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(01):106–109, 1952.
- [4] Andrew Blake and Andrew Zisserman. *Visual reconstruction*. MIT press Cambridge, 1987.
- [5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

- [6] Luigi Ambrosio and Vincenzo Maria Tortorelli. Approximation of functional depending on jumps by elliptic functional via Γ -convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.
- [7] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM Journal on Applied Mathematics*, 55(3):827–863, 1995.
- [8] O. Wittich, A. Kempe, G. Winkler, and V. Liebscher. Complexity penalized least squares estimators: Analytical results. *Mathematische Nachrichten*, 281(4):582–595, 2008.
- [9] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37(1):157–183, 2009.
- [10] M. Fornasier and R. Ward. Iterative thresholding meets free-discontinuity problems. *Foundations of Computational Mathematics*, 10(5):527–567, 2010.
- [11] M. Fornasier, R. March, and F. Solombrino. Existence of minimizers of the Mumford-Shah functional with singular operators and unbounded data. *Annali di Matematica Pura ed Applicata*, 192(3):361–391, 2013.
- [12] M. Jiang, P. Maass, and T. Page. Regularizing properties of the Mumford-Shah functional for imaging applications. *Inverse Problems*, 30(3):035007, 2014.
- [13] O. Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Cornell University, 1999.
- [14] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [15] B. Alexeev and R. Ward. On the complexity of Mumford–Shah-type regularization, viewed as a relaxed sparsity constraint. *IEEE Transactions on Image Processing*, 19(10):2787–2789, 2010.
- [16] Andy Tsai, Anthony Yezzi Jr, and A. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10(8):1169–1186, 2001.
- [17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1133–1140, 2009.
- [18] D. Massonnet and K. Feigl. Radar interferometry and its application to changes in the earth’s surface. *Reviews of Geophysics*, 36:441–500, 1998.
- [19] T. Chan, S. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color models. *Journal of Visual Communication and Image Representation*, 12:422–435, 2001.

- [20] L. Vese and S. Osher. Numerical methods for p-harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis*, 40:2085–2104, 2002.
- [21] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine. *Journal of Visual Communication and Image Representation*, 13:238–248, 2002.
- [22] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.
- [23] Inam Ur Rahman, Iddo Drori, Victoria C Stodden, David L Donoho, and Peter Schröder. Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation*, 4(4):1201–1232, 2005.
- [24] G. Rosman, M. Bronstein, A. Bronstein, A. Wolf, and R. Kimmel. Group-valued regularization framework for motion segmentation of dynamic non-rigid shapes. In *Scale Space and Variational Methods in Computer Vision*, pages 725–736. Springer, 2012.
- [25] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Applied and Computational Harmonic Analysis*, 27:325–333, 2009.
- [26] A. Weinmann. Interpolatory multiscale representation for functions between manifolds. *SIAM Journal on Mathematical Analysis*, 44:162–191, 2012.
- [27] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages I948–I953, 2001.
- [28] Christophe Chef d’Hotel, David Tschumperlé, Rachid Deriche, and O Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.
- [29] Philipp Grohs, Hanne Hardering, and Oliver Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Foundations of Computational Mathematics (to appear)*, 2014.
- [30] P. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105:171–185, 2013.
- [31] P Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.
- [32] P. Fletcher, C. Lu, S. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23:995–1005, 2004.
- [33] J. Oller and J. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, pages 1562–1581, 1995.
- [34] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds I. *Annals of Statistics*, pages 1–29, 2003.
- [35] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds II. *Annals of Statistics*, pages 1225–1259, 2005.

- [36] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
- [37] Peter Basser, James Mattiello, and Denis LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.
- [38] H. Johansen-Berg and T. Behrens. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*. Academic Press, London, 2009.
- [39] Yaniv Assaf and Ofer Pasternak. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *Journal of Molecular Neuroscience*, 34(1):51–61, 2008.
- [40] J Foong, M Maier, CA Clark, GJ Barker, DH Miller, and MA Ron. Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(2):242–244, 2000.
- [41] Marek Kubicki, Robert McCarley, Carl-Fredrik Westin, Hae-Jeong Park, Stephan Maier, Ron Kikinis, Ferenc Jolesz, and Martha Shenton. A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, 41(1):15–30, 2007.
- [42] Andrew Alexander, Jee Lee, Mariana Lazar, Rebecca Boudos, Molly DuBray, Terrence Oakes, Judith Miller, Jeffrey Lu, Eun-Kee Jeong, William McMahon, et al. Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage*, 34(1):61–73, 2007.
- [43] H Rosas, Stephanie Lee, Alexander Bender, Alexandra Zaleta, Mark Vangel, Peng Yu, Bruce Fischl, Vasanth Pappu, Christina Onorato, Jang-Ho Cha, et al. Altered white matter microstructure in the corpus callosum in Huntington’s disease: implications for cortical “disconnection”. *Neuroimage*, 49(4):2995–3004, 2010.
- [44] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.
- [45] C Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.
- [46] D Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372, 2004.
- [47] Maxime Descoteaux, Elaine Angelino, Shaun Fitzgibbons, and Rachid Deriche. Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3):497–510, 2007.
- [48] Christopher Hess, Pratik Mukherjee, Eric Han, Duan Xu, and Daniel Vigneron. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magnetic Resonance in Medicine*, 56(1):104–117, 2006.

- [49] T Behrens, H Johansen-Berg, Saad Jbabdi, M Rushworth, and M Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155, 2007.
- [50] Alvina Goh, Christophe Lenglet, P Thompson, and René Vidal. A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI). In *IEEE Conference on Computer Vision and Pattern Recognition.*, pages 2496–2503, 2009.
- [51] Zhizhou Wang and Baba Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Transactions on Medical Imaging*, 24(10):1267–1277, 2005.
- [52] Zhizhou Wang and Baba Vemuri. An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition.*, pages 1228–1233, 2004.
- [53] Guang Cheng, Hesamoddin Salehian, and Baba Vemuri. Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation. In *Computer Vision–ECCV 2012*, pages 390–401. Springer, 2012.
- [54] M Wiegell, D Tuch, H Larsson, and V Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *NeuroImage*, 19(2):391–401, 2003.
- [55] Lisa Jonasson, Xavier Bresson, Patric Hagmann, Olivier Cuisenaire, Reto Meuli, and Jean-Philippe Thiran. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis*, 9(3):223–236, 2005.
- [56] Christian Feddern, Joachim Weickert, and Bernhard Burgeth. Level-set methods for tensor-valued images. In *Proc. Second IEEE Workshop on Geometric and Level Set Methods in Computer Vision*, pages 65–72, 2003.
- [57] Leonid Zhukov, Ross Whitaker, Ken Museth, David Breen, and Alan H Barr. Level set modeling and segmentation of diffusion tensor magnetic resonance imaging brain data. *Journal of Electronic Imaging*, 12(1):125–133, 2003.
- [58] Bernd Kirchheim. Rectifiable metric spaces: local structure and regularity of the Hausdorff measure. *Proceedings of the American Mathematical Society*, 121(1):113–123, 1994.
- [59] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *Journal of Nonparametric Statistics*, 14(1-2):203–222, 2002.
- [60] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity penalized M-estimation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.
- [61] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences (to appear)*, *arXiv:1312.7710*, 2014.
- [62] A. Weinmann, M. Storath, and L. Demaret. The L^1 -Potts functional for robust jump-sparse reconstruction. 2014.

- [63] M. Storath, A. Weinmann, and L. Demaret. Jump-sparse and sparse recovery using Potts functionals. *IEEE Transactions on Signal Processing*, 62(14):3654–3666, 2014.
- [64] R. Killick, P. Fearnhead, and I. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.
- [65] M. Storath and A. Weinmann. Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences*, 7(3):1826–1852, 2014.
- [66] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.
- [67] W. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3:371–406, 1990.
- [68] Bijan Afsari, Roberto Tron, and René Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.
- [69] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso. Newton algorithms for Riemannian distance related problems on connected locally symmetric manifolds. *IEEE Journal of Selected Topics in Signal Processing*, 7:634–645, 2013.
- [70] Marc Arnaudon and Frank Nielsen. On approximating the Riemannian 1-center. *Computational Geometry*, 46(1):93–104, 2013.
- [71] J. Wallner and N. Dyn. Convergence and C^1 analysis of subdivision schemes on manifolds by proximity. *Computer Aided Geometric Design*, 22:593–622, 2005.
- [72] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization (to appear)*, 2014.
- [73] M. do Carmo. *Riemannian Geometry*. Birkhäuser, Boston, 1992.
- [74] W. Ballmann, M. Gromov, and V. Schroeder. *Manifolds of nonpositive curvature*. Birkhäuser, Boston, 1985.
- [75] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces*, volume 338 of *Contemp. Math.*, pages 357–390. American Mathematical Society, Providence, 2003.
- [76] A. Chambolle. Finite-differences discretizations of the Mumford-Shah functional. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(02):261–288, 1999.
- [77] M. Storath, A. Weinmann, J. Friel, and M. Unser. Joint image reconstruction and segmentation using the Potts model. *submitted, preprint arXiv:1405.5850*, 2014.
- [78] D. Bertsekas. Multiplier methods: a survey. *Automatica*, 12(2):133–145, 1976.

- [79] D. Le Bihan, J.-F. Mangin, C. Poupon, C. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001.
- [80] B. Chen and E. Hsu. Noise removal in magnetic resonance diffusion tensor imaging. *Magnetic Resonance in Medicine*, 54:393–401, 2005.
- [81] S. Basu, T. Fletcher, and R. Whitaker. Rician noise removal in diffusion tensor MRI. In *Medical Image Computing and Computer-Assisted Intervention 2006*, pages 117–125. Springer, 2006.
- [82] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pages 115–122. Springer, 2005.
- [83] P. Cook, Y. Bai, S. Nedjati-Gilani, K. Seunarine, M. Hall, G. Parker, and D. Alexander. Camino: Open-source diffusion-MRI reconstruction and processing. In *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 2759, 2006.
- [84] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache. Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–1482, 2007.
- [85] DC Alexander, GJ Barker, and SR Arridge. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine*, 48(2):331–340, 2002.
- [86] D Tuch, T Reese, M Wiegell, N Makris, J Belliveau, and V Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48(4):577–582, 2002.
- [87] Lawrence Frank. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 47(6):1083–1099, 2002.
- [88] Evren Özarslan and Thomas Mareci. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 50(5):955–965, 2003.